

THE AMENDMENTS

In The Claims

1. (Currently Amended) A system for allocating a resource in a network comprising: multiple network data ports to connect to the network to receive and route requests to servers in the network;

first logic coupled to the network data ports for determining if any of a plurality of persistence policies, comprising at least one first persistence policy and at least one second persistence policy, is applicable to a service request from a client in the network, and, if so, allocating the resource to the request based on application of the persistence policy determined to be applicable, wherein the first persistence policy, if applicable, directs the service request to a server, previously or currently connected to the client, that is identified based on content of the service request, the content comprising server, session or cookie information within the service request or corresponding packet at one or more layers corresponding to OSI layers 5-7, and wherein the second persistence policy, if applicable, directs the service request to a server, previously or currently connected to the client, that is identified based on client information within the service request or corresponding packet;

the first logic configured so that:

when a service request for a content-enabled service, comprising a request for a connection to a server at a layer corresponding to any of OSI layers 5-7, is received, the first logic determines if the at least one first persistence policy is applicable;

when a service request for a non-content-enabled service, comprising a request for a connection to a server at a layer corresponding to OSI layer 4, is received, the first logic determines if the at least one second persistence policy is applicable;

when the service request for a content-enabled service is received but it is determined that the at least one first persistence policy is inapplicable, the first logic determines if the at least one second persistence policy is applicable;

wherein the first logic is shared by and supports service requests for both content-enabled and non-content enabled services; and

second logic coupled to the network data ports for allocating the resource to the request based on application of a load balancing policy only if none of the plurality of persistence policies is determined to be applicable as determined by the first logic;

wherein, when the service request for a content-enabled service is bound to a server identified through application of a load balancing policy, such request is bound to the server only after multiple persistence policies are checked for applicability.

2. (Original) The system of claim 1 wherein the first logic determines if a persistence policy is applicable to a service request through consideration of whether or not an allocation exists or recently expired for an originator the service request.

3. (Currently Amended) A system for allocating a resource, in a network, to a resource request, the request having an originator, based on application of a persistence policy comprising:

multiple network data ports to connect to the network to receive and route requests to servers in the network;

first logic coupled to the data ports for determining whether an allocation exists or recently expired for the originator of the resource request, based on application of any of a plurality of persistence policies, comprising at least one first persistence policy and at least one second persistence policy, to the request, and, if so, identifying the resource which is the subject of the existing or recently expired allocation, wherein the first persistence policy, if applicable, directs the resource request to a server, currently or previously connected to the originator of the request or corresponding client, that is identified based on content of the resource request, the content comprising server, session or cookie information within the resource request or corresponding packet at one or more layers corresponding to OSI layers 5-7, and wherein the second persistence policy, if applicable, directs the resource request to a server, currently or previously connected to the originator of the request or corresponding client, that is identified based on client information within the request or corresponding packet;

the first logic configured so that:

when a resource request for a content-enabled service, comprising a request for a connection to a server at a layer corresponding to any of OSI layers 5-7, is received, the first logic determines whether an allocation exists or recently expired for the originator of the resource request by applying the at least one first persistence policy, or the at least one second persistence policy when the at least one first persistence policy is determined to be inapplicable;

when a resource request for a non-content-enabled service, comprising a request for a connection to a server at a layer corresponding to OSI layer 4, is received, the first logic determines whether an allocation exists or recently expired for the originator of the resource request by applying the at least one second persistence policy;

wherein the first logic is shared by and supports resource requests for both content enabled and non-content enabled services; and
second logic in the network for allocating the resource, once identified, to the resource request;

wherein, when the service request for a content-enabled service is bound to a server identified through application of a load balancing policy, such request is bound to the server only after multiple persistence policies are checked for applicability.

4. (Original) The system of claim 3 wherein the resource request is derived from or represented by a packet.

5. (Withdrawn) A system for maintaining a data structure useful for allocating a resource to a resource request based on application of a persistence policy comprising:

first logic for making an entry in the data structure representing an allocation, upon or after implementation of the allocation, and time-stamping the entry with a time-stamp indicating the time when or about when the allocation is terminated; and

second logic for scanning the data structure, and deleting entries for which a time-out condition is determined to exist.

6. (Withdrawn) The system of claim 5 wherein the second logic has access to a current time, and determines that a time-out condition exists if the time-stamp value equals or exceeds the current time by a predetermined amount.

7. (Withdrawn) The system of claim 5 wherein the second logic repetitively scans the data structure.

8. (Withdrawn) The system of claim 5 wherein the second logic periodically scans the data structure.

9. (Withdrawn) The system of claim 6 wherein the predetermined amount is programmable.

10. (Withdrawn) A system for making an entry in a data structure representing an allocation, the data structure being useful for allocating a resource to a resource request based on application of a persistence policy, the system comprising:

first logic for deriving a first index from information relating to the resource request if such information is available, and using the first index to make an entry in the data structure representing the allocation if the first index is available; and

second logic for deriving a second index from information relating to the resource request, and using the second index to make an entry in the data structure representing the allocation.

11. (Withdrawn) The system of claim 10 wherein the data structure is a history table.

12. (Withdrawn) The system of claim 10 wherein the first logic derives each of the first and second indices by applying a hashing function to information derived from the resource request.

13. (Withdrawn) The system of claim 10 wherein the first logic derives the first index by applying a hashing function to a hashing key derived from a session or cookie identifier derived from a packet spawning the resource request.

14. (Withdrawn) The system of claim 10 where the second logic derives the second index by applying a hashing function to a hashing key derived from a client IP address derived from a packet spawning the resource request.

15. (Withdrawn) A system for making an entry in a data structure representing an allocation, the data structure being useful for allocating a resource to a resource request based on application of a persistence policy, the system comprising:

first means for deriving a first index from information relating to the resource request if such information is available, and using the first index to make an entry in the data structure representing the allocation if the first index is available; and

second means for deriving a second index from information relating to the resource request, and using the second index to make an entry in the data structure representing the allocation.

16. (Withdrawn) A system for accessing a data structure in order to allocate a resource to a resource request based on application of a persistence policy, entries in the data structure corresponding to allocated resources, the system comprising:

first logic for deriving a first index from information relating to a resource request if such information is available, using the first index to access the data structure and determine if an entry corresponding to the first index is available, and, if such an entry is available, allocating the resource corresponding to the entry to the resource request; and

second logic for deriving, if the first index or an entry corresponding to the first index is unavailable, a second index from information relating to the resource request, and using the second index to access the data structure and determine if an entry corresponding to the second index is available, and, if such an entry is available, allocating the resource corresponding to the entry to the resource request.

17. (Withdrawn) The system of claim 16 further comprising third logic for allocating, if an entry corresponding to the second index is unavailable, a resource to the request based on application of a load balancing policy or other persistence policy.

18. (Withdrawn) The system of claim 17 further comprising fourth means for using the first index to make an entry in the data structure corresponding to the allocation of claim 17 if such first index is available.

19. (Withdrawn) The system of claim 18 further comprising fifth means for using the second index to make an entry in the data structure corresponding to the allocation of claim 17.

20. (Currently Amended) A method of allocating a resource in a network comprising: providing multiple data ports to connect to the network to receive and route requests to servers in the network;

determining if any of a plurality of persistence policies, comprising at least one first persistence policy and at least one second persistence policy, is applicable to a service request from a client in the network, and, if so, allocating the resource to the request based on application of the persistence policy determined to be applicable, wherein the first persistence policy, if applicable, directs the service request to a server, currently or previously connected to the client, that is identified based on content of the service request, the content comprising server, session or cookie information within the service request or corresponding packet at one or more layers corresponding to OSI layers 5-7, and wherein the second persistence policy, if applicable, directs the service request to a server, currently or previously connected to the client, that is identified based on client information within the service request or corresponding packet;

when a service request for a content-enabled service, comprising a request for a connection to a server at a layer corresponding to any of OSI layers 5-7, is received, the determining step comprises determining if the at least one first persistence policy is applicable;

when a service request for a non-content-enabled service, comprising a request for a connection to a server at a layer corresponding to OSI layer 4, is received, the determining step comprises determining if the at least one second persistence policy is applicable;

when the service request for a content-enabled service is received, but it is determined that the at least one first persistence policy is inapplicable, the determining step comprises determining if the at least one second persistence policy is applicable;

performing the foregoing steps through logic coupled to the data ports that is shared by and supports service requests for both content enabled and non-content enabled services; and

allocating the resource to the request based on application of a load balancing policy only if none of the persistence policies in the plurality of persistence policies is determined to be applicable in the foregoing determining step;

wherein, when the service request for a content-enabled service is bound to a server identified through application of a load balancing policy, such request is bound to the server only after multiple persistence policies are checked for applicability.

21. (Currently Amended) A method of allocating a resource, in a network, to a resource request based on application of a persistence policy, the request having an originator, comprising:

providing multiple network data ports to connect to the network to receive and route requests to servers in the network;

determining whether an allocation exists or recently expired for the originator of the resource request, based on application of any of a plurality of persistence policies, comprising at least one first persistence policy and at least one second persistence policy, to the request, and, if so, identifying the resource which is the subject of the existing or recently expired allocation, wherein the first persistence policy, if applicable, directs the resource request to a server, currently or previously connected to the originator of the resource request or corresponding client, that is identified based on content of the resource request, the content comprising server, session or cookie information within the resource request or corresponding packet at one or more layers corresponding to OSI layers 5-7, and wherein the second persistence policy, if applicable, directs the resource request to a server, currently or previously connected to the originator of the request or corresponding client, that is identified based on information within the resource request or corresponding packet other than content;

when a resource request for a content-enabled service, comprising a request for a connection to a server at a layer corresponding to any of OSI layers 5-7, is received, the determining step comprises determining whether an allocation exists or recently expired for the originator of the resource request by applying the at least one first persistence policy, or the least one second persistence policy when the at least one first persistence policy is determined to be inapplicable;

when a resource request for a non-content-enabled service, comprising a request for a connection to a server at a layer corresponding to OSI layer 4, is received, the determining step comprises determining whether an allocation exists or recently expired for the originator of the request by applying the at least one second persistence policy;

performing the foregoing steps through logic coupled to the data ports that is shared by and supports both resource requests for content enabled and non-content enabled services; and

allocating the resource in the network, once identified, to the resource request;

wherein, when the service request for a content-enabled service is bound to a server identified through application of a load balancing policy, such request is bound to the server only after multiple persistence policies are checked for applicability.

22. (Original) The method of claim 21 wherein the resource request is spawned by a packet.

23. (Withdrawn) A method of maintaining a data structure useful for allocating a resource to a resource request based on application of a persistence policy comprising:
making an entry in the data structure representing an allocation, and time-stamping the entry with a time-stamp indicating the time when or about when the allocation is terminated; and
scanning the data structure, and deleting entries for which a time-out condition is determined to exist.

24. (Withdrawn) The method of claim 23 further comprising determining that a time-out condition exists if the time-stamp value equals or exceeds a current time by a predetermined amount.

25. (Withdrawn) The method of claim 23 further comprising repetitively scanning the data structure.

26. (Withdrawn) The method of claim 23 further comprising periodically scanning the data structure.

27. (Withdrawn) The method of claim 24 wherein the predetermined amount is programmable.

28. (Withdrawn) A method of making an entry in a data structure representing an allocation, the data structure being useful for allocating a resource to a resource request based on application of a persistence policy, the method comprising:

deriving a first index from information relating to the resource request if such information is available;

using the first index to make an entry in the data structure representing the allocation if the first index is available;

deriving a second index from information relating to the resource request; and

using the second index to make an entry in the data structure representing the allocation.

29. (Withdrawn) The method of claim 28 wherein the data structure is a history table.

30. (Withdrawn) The method of claim 28 further comprising deriving each of the first and second indices by applying a hashing function to information derived from a packet spawning the resource request.

31. (Withdrawn) The method of claim 28 further comprising deriving the first index by applying a hashing function to a hashing key derived from a session or cookie identifier derived in turn from a packet spawning the resource request.

32. (Withdrawn) The method of claim 28 further comprising deriving the second index by applying a hashing function to a hashing key derived from a client IP address derived in turn from a packet spawning the resource request.

33. (Withdrawn) A method of making an entry in a data structure representing an allocation, the data structure being useful for allocating a resource to a resource request based on application of a persistence policy, the method comprising:

a step for deriving a first index from information relating to the resource request if such information is available;

a step for using the first index to make an entry in the data structure representing the allocation if the first index is available;

a step for deriving a second index from information relating to the resource request; and

a step for using the second index to make an entry in the data structure representing the allocation.

34. (Withdrawn) A method of accessing a data structure in order to allocate a resource to a resource request based on application of a persistence policy, entries in the data structure corresponding to allocated resources, the method comprising:

deriving a first index from information relating to a resource request if such information is available;

using the first index to access the data structure and determine if an entry corresponding to the first index is available;

if such an entry is available, allocating the resource corresponding to the entry to the resource request;

deriving, if the first index or an entry corresponding to the first index is unavailable, a second index from information relating to the resource request;

using the second index to access the data structure and determine if an entry corresponding to the second index is available; and

if such an entry is available, allocating the resource corresponding to the entry to the resource request.

35. (Withdrawn) The method of claim 34 further comprising allocating, if an entry corresponding to the second index is unavailable, a resource to the request based on application of a load balancing policy or other persistence policy.

36. (Withdrawn) The method of claim 35 further comprising using the first index to make an entry in the data structure corresponding to the allocation of claim 35 if such first index is available.

37. (Withdrawn) The method of claim 36 further comprising using the second index to make an entry in the data structure corresponding to the allocation of claim 35.

38. (Withdrawn) A method of accessing a data structure in order to allocate a resource to a resource request based on application of a persistence policy, entries in the data structure corresponding to allocated resources, the method comprising:

- a step for deriving a first index from information relating to a resource request if such information is available;

- a step for using the first index to access the data structure and determine if an entry corresponding to the first index is available;

- a step for allocating, if such an entry is available, the resource corresponding to the entry to the resource request;

- a step for deriving, if the first index or an entry corresponding to the first index is unavailable, a second index from information relating to the resource request;

- a step for using the second index to access the data structure and determine if an entry corresponding to the second index is available; and

- a step for allocating, if such an entry is available, the resource corresponding to the entry to the resource request.

39. (Currently Amended) The system of claim 1 wherein the at least one first persistence policy comprises ~~at least one~~ or more of a cookie-based hashing stickiness persistence policy, a session-ID hashing stickiness persistence policy, and a self-identification stickiness persistence policy, and the at least one second persistence policy comprises at least one client-based persistence policy.

40. (Previously Presented) The system of claim 1 wherein the at least one first persistence policy comprises at least one session-based persistence policy, and the at least one second persistence policy comprises at least one client-based persistence policy.

41. (Currently Amended) A system for allocating a resource in a network comprising: multiple network data ports to connect to the network to receive and route requests to servers in the network;

first logic coupled to the network data ports for determining if any of a plurality of persistence policies, comprising at least one cookie-based persistence policy, at least one session-based persistence policy, and at least one client-based persistence policy, is applicable to a service request from a client in the network, and, if so, allocating the resource to the request based on application of the persistence policy determined to be applicable, wherein the first persistence policy, if applicable, directs the service request to a server, currently or previously connected to the client, that is identified based on content of the service request, the content comprising server, session or cookie information within the service request or corresponding packet at one or more layers corresponding to OSI layers 5-7, and wherein the second persistence policy, if applicable, directs the service request to a server, currently or previously connected to the client, that is identified based on client information within the service request or corresponding packet at a layer corresponding to OSI layer 4;

the first logic configured so that:

when a service request for a content-enabled service, comprising a request for a connection to a server at a layer corresponding to any of OSI layers 5-7, is received, the first logic determines if either the at least one cookie-based or at least one session-based persistence policy is applicable;

when a service request for a non-content-enabled service, comprising a request for a connection to a server at a layer corresponding to OSI layer 4, is received, the first logic determines if the at least one client-based persistence policy is applicable;

when the service request for a content-enabled service is received but it is determined that both the at least one cookie-based persistence policy and the at least one session-based persistence policy are inapplicable, the first logic determines if the at least one client-based persistence policy is applicable;

wherein the first logic is shared by and supports service requests for both content-enabled and non-content enabled services; and

second logic in the network for allocating the resource to the request based on application of a load balancing policy only if none of the plurality of persistence policies is determined to be applicable as determined by the first logic;

wherein, when the service request for a content-enabled service is bound to a server identified through application of a load balancing policy, such request is bound to the server only after multiple persistence policies are checked for applicability.

42. (Canceled).

43. (Previously Presented) The system of claim 1 where the at least one first persistence policy is content-aware, and the at least one second persistence policy is non-content aware.